Phase 6.2 — Heteroscedastic GP Surrogate

Feature map

Let $\mathcal{P}$ denote the space of PACKAGE configurations. A package $p \in \mathcal{P}$ is encoded as a fixed-length feature vector

$$\boldsymbol{x} = \varphi(p) \in \mathbb{R}^d$$

where φ concatenates one-hot encodings of the model slot, a binary skill-presence vector, and one-hot encodings of the named system-prompt and template-value variants (the Phase 3.1 slot-space schema).

Observations

After N completed trials we have a dataset $\mathcal{D}_N = \{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^N$ where each $\boldsymbol{y}_i \in \mathbb{R}^5$ records the five Pareto-axis values

$$\boldsymbol{y}_i = \big(\bar{\tau}_i,\ \bar{c}_i,\ s_i,\ \bar{q}_i,\ r_i\big) \qquad \text{(tokens, dollars, scaling slope, quality, subjective)}$$

Trials without a subjective score contribute only the first four components; the subjective head is trained on the subset of fully-scored trials.

Heteroscedastic single-task GP (per output head)

For each output $j \in \{1, \ldots, 5\}$ we place an independent GP prior:

$$f^{(j)}(\boldsymbol{x}) \sim \mathcal{GP}\Big(0,\ k^{(j)}(\boldsymbol{x}, \boldsymbol{x}')\Big)$$

with a squared-exponential (RBF-ARD) kernel

$$k^{(j)}(\boldsymbol{x}, \boldsymbol{x}') = \sigma_{f,j}^2 \exp\big(-\tfrac{1}{2}(\boldsymbol{x} - \boldsymbol{x}')^\top \boldsymbol{\Lambda}_j^{-1}(\boldsymbol{x} - \boldsymbol{x}')\big)$$

where $\boldsymbol{\Lambda}_j = \operatorname{diag}(\ell_{j,1}^2, \ldots, \ell_{j,d}^2)$ are per-dimension length-scales (automatic relevance determination).

Input-dependent noise. Rather than assuming a fixed noise level, we model the log noise variance with a second GP:

$$g^{(j)}(\boldsymbol{x}) \sim \mathcal{GP}\Big(0,\ k_\varepsilon^{(j)}(\boldsymbol{x}, \boldsymbol{x}')\Big) \qquad \sigma^{2,(j)}(\boldsymbol{x}) = \exp\Big(g^{(j)}(\boldsymbol{x})\Big)$$

The observation model is then

$$y_i^{(j)} = f^{(j)}(\boldsymbol{x}_i) + \varepsilon_i^{(j)}, \qquad \varepsilon_i^{(j)} \sim \mathcal{N}\Big(0,\ \sigma^{2,(j)}(\boldsymbol{x}_i)\Big)$$

Posterior

Let $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]^\top$, $\boldsymbol{y}^{(j)} = [y_1^{(j)}, \ldots, y_N^{(j)}]^\top$, and $\boldsymbol{\Sigma}^{(j)} = \mathrm{diag}\big(\sigma^{2,(j)}(\boldsymbol{x}_1), \ldots, \sigma^{2,(j)}(\boldsymbol{x}_N)\big)$.

The posterior over $f^{(j)}$ at a new point $\boldsymbol{x}_*$ is Gaussian:

$$\mu_N^{(j)}(\boldsymbol{x}_*) = \boldsymbol{k}_*^{(j)\top} \left[\boldsymbol{K}^{(j)} + \boldsymbol{\Sigma}^{(j)}\right]^{-1} \boldsymbol{y}^{(j)} \tag{1}$$

$$\sigma_N^{2,(j)}(\boldsymbol{x}_*) = k^{(j)}(\boldsymbol{x}_*, \boldsymbol{x}_*) - \boldsymbol{k}_*^{(j)\top} \left[\boldsymbol{K}^{(j)} + \boldsymbol{\Sigma}^{(j)}\right]^{-1} \boldsymbol{k}_*^{(j)} \tag{2}$$

where $\boldsymbol{K}_{ij}^{(j)} = k^{(j)}(\boldsymbol{x}_i, \boldsymbol{x}_j)$ and $\boldsymbol{k}_*^{(j)} = [k^{(j)}(\boldsymbol{x}_*, \boldsymbol{x}_i)]_{i=1}^N$.

Hyperparameters $\theta_j = \{\sigma_{f,j}^2,\, \boldsymbol{\Lambda}_j,\, \theta_{\varepsilon,j}\}$ are optimised by maximising the log marginal likelihood.

Bootstrap discipline

Define the trial count threshold $N_0 \approx 10$. The proposer operates in two regimes:

$$\text{next proposal} = \begin{cases} \text{RandomFromSlotSpace} & N < N_0 \\ \arg\max_{\boldsymbol{x} \in \mathcal{X}}\ \alpha_{\text{EHVI}}(\boldsymbol{x};\, \mathcal{D}_N) & N \geq N_0 \end{cases}$$

where $\mathcal{X}$ is the discrete set of un-evaluated packages and α_{EHVI} is the expected hypervolume improvement over the current 5D Pareto frontier (Phase 6.3).